

A Low-Complexity Recursive Approach Toward Code-Domain NOMA for Massive Communications

Mohammad Vahid Jamali, and Hessam Mahdaviifar

Electrical Engineering and Computer Science Department, University of Michigan, Ann Arbor, MI 48109, USA

Email: mvjamali@umich.edu, hessam@umich.edu

Abstract—Nonorthogonal multiple access (NOMA) is a promising technology to meet the demands of the next generation wireless networks on massive connectivity, high throughput and reliability, improved fairness, and low latency. In this context, code-domain NOMA which attempts to serve K users in $M \leq K$ orthogonal resource blocks, using a pattern matrix, is of utmost interest. However, extending the pattern matrix dimensions severely increases the detection complexity and hampers on the significant advantages that can be achieved using large pattern matrices. In this paper, we propose a novel approach toward code-domain NOMA which factorizes the pattern matrix as the Kronecker product of some other factor matrices each with a smaller dimension. Therefore, both the pattern matrix design at the transmitter side and the mixed symbols' detection at the receiver side can be performed over much smaller dimensions and with a remarkably reduced complexity and latency. As a consequence, the system can significantly be overloaded to effectively support the requirements of the next generation wireless networks without any considerable increase on the system complexity.

I. INTRODUCTION

Nonorthogonal multiple access (NOMA) is becoming one of the key enabling technologies for the fifth-generation (5G) wireless networks and the applications beyond, including the Internet of Things (IoT), and massive machine-type communications (MMTC) [1], [2]. NOMA, through serving multiple users in the same orthogonal resource block, or equivalently, a set of users in a smaller set of resource elements (REs), attempts to significantly increase the system throughput and reliability, improve the users' fairness, reduce the latency, and support the massive connectivity [3]. NOMA is a general principle and many of the recently proposed multiple access techniques can be considered as special cases. However, generally speaking, NOMA techniques can be classified into two different categories: power-domain NOMA, and code-domain NOMA.

Power-domain NOMA serves multiple users in the same orthogonal resource block through allocating different power levels to different users. In contrast to the traditional power allocation methods, the users with poorer channel conditions are assigned more powers [4]. Power-domain NOMA has attracted significant attention in recent years and many interesting problems have been addressed in this context, including cooperative NOMA [5], NOMA with simultaneous wireless information and power transfer (SWIPT) [6], multiple-input multiple-output (MIMO) NOMA [7], and NOMA in millimeter wave communications [8]. Detection procedure in power-domain NOMA highly relies on successive interference cancellation (SIC). Therefore, when the channel condition of different users paired together is close to each other, SIC-based power-domain NOMA does not work well. In order to enable (near) power-balanced NOMA, the authors of [9] proposed network-coded multiple access (NOMA); a new NOMA architecture which jointly employs the physical-layer network coding and multiuser decoding to significantly increase the NOMA throughput.

Code-domain NOMA, on the other hand, attempts to serve a set of users, say K , in a set of M orthogonal resource blocks,

$M \leq K$, using a pattern matrix. The pattern matrix comprises K pattern vectors assigned to each user specifying the set of REs available to them. As an interesting feature of code-domain NOMA, it can work well even in power-balanced scenarios when each user is served by a unique pattern vector. Nevertheless, code-domain NOMA has received relatively less attention compared to the power-domain counterpart, because it usually requires rather complex multiuser nonlinear detection algorithms, such as message passing algorithm (MPA), maximum likelihood (ML) detection, and maximum *a posteriori* (MAP) detection. In this context, sparse code multiple access (SCMA) was proposed in [10] based on directly mapping the incoming bits to multidimensional codewords of a SCMA codebook set. A low complexity SCMA decoding algorithm has been further suggested in [11] based on list sphere decoding. Moreover, lattice partition multiple access (LPMA) has been proposed in [12], [13] based on multilevel lattice codes for multiple users. Interleave-grid multiple access (IGMA) has also been proposed in order to increase the user multiplexing capability and improve the performance [14]. In addition, pattern division multiple access (PDMA) has been introduced in [15], [16], where the pattern vectors are designed with disparate order of transmission diversity to mitigate the error propagation problem in SIC receivers. As a consequence, the sparsity feature of the pattern matrix does not necessarily hold, i.e., the number of REs allocated to a particular user may be comparable to the total number of resource blocks M .

Design of the pattern matrix plays a critical role in code-domain NOMA to balance the trade-off between the system performance and complexity. Impact of pattern matrix on the average sum rate of SCMA systems has been investigated in [17] where a low-complexity iterative algorithm has been proposed to facilitate the design of pattern matrix. Moreover, the total throughput of low-density code-domain (LDCD) NOMA has recently been characterized in [18] for regular random pattern matrices with large dimensions. It is well understood that, for a given overload factor $\beta = K/M$, the higher the dimension of pattern matrix we have, the better the performance we get. This is because more number of REs can be assigned to each user for a higher dimension of pattern matrix [16]. However, increasing the pattern matrix dimension results in a significantly higher detection complexity. Inspired by this trade-off between the system performance and detection complexity, we propose a novel approach toward code-domain NOMA which factorizes the pattern matrix as the Kronecker product of some other smaller factor matrices. As a consequence, both the pattern matrix design at the transmitter and the mixed symbols' detection at the receiver can be performed over much smaller dimensions and with a remarkably reduced complexity, through recursive detection; hence, the system performance can significantly be improved without any considerable increase on the complexity.

The rest of the paper is organized as follows. In Section II, we briefly review the basics of code-domain NOMA with pattern matrix and highlight the preliminaries. In Section III, we

describe the proposed method, and provide an illustrating example. In Section IV, we classify the design procedures, obtain the sum rate formula, and analyze the detection complexity. We provide the numerical results in Section V, and conclude the paper in Section VI.

II. PRELIMINARIES AND SYSTEM MODEL

We consider a collection of K users communicating using M REs, and define the overload factor as $\beta = K/M \geq 1$ implying the multiplexing gain of PDMA compared to orthogonal multiple access (OMA). The modulation symbol x_k of the k -th user is spread over M orthogonal REs using the pattern vector \mathbf{g}_k as $\mathbf{v}_k = \mathbf{g}_k x_k$, $1 \leq k \leq K$, where $\mathbf{g}_k \in \mathcal{B}^{M \times 1}$, with $\mathcal{B} = \{0, 1\}$, is an $M \times 1$ binary vector defining the set of REs available to the k -th user; the k -th user can use the i -th RE, $1 \leq i \leq M$, if the i -th element of \mathbf{g}_k is "1", i.e., $g_{i,k} = \mathbf{g}_k(i) = 1$, and vice versa. Then the $(M \times K)$ -dimensional pattern matrix $\mathbf{G} \in \mathcal{B}^{M \times K}$ can be constructed as [16]

$$\mathbf{G}_{M \times K} = [\mathbf{g}_1 \quad \mathbf{g}_2 \quad \cdots \quad \mathbf{g}_K] = [g_{i,k}]_{M \times K}. \quad (1)$$

For the uplink transmission each k -th user transmits the spread symbol, using its pattern vector, to the base station (BS). Therefore, assuming perfect synchronization at the BS, the received $(M \times 1)$ -dimensional vector, comprising the received signal at each of M REs, can be modeled as the Gaussian vector channel model of

$$\mathbf{y} = \sum_{k=1}^K \text{diag}(\mathbf{h}_k) \mathbf{v}_k + \mathbf{n}, \quad (2)$$

where \mathbf{h}_k is the vector modeling the uplink channel response of the k -th user at all of M REs, and \mathbf{n} is the noise vector at the BS with length M . Furthermore, $\text{diag}(\mathbf{h}_k)$ is a diagonal matrix consisting of the elements of \mathbf{h}_k . Then the uplink transmission model can be reformulated as

$$\mathbf{y} = \mathbf{H} \mathbf{x} + \mathbf{n}, \quad (3)$$

in which $\mathbf{x} = [x_1, x_2, \dots, x_K]^T$, $\mathbf{H} = \mathcal{H} \bullet \mathbf{G}_{M \times K}$ is the PDMA equivalent uplink channel response, $\mathcal{H} = [\mathbf{h}_1, \mathbf{h}_2, \dots, \mathbf{h}_K]$, and \bullet denotes the element-wise product [16].

Moreover, for the downlink transmission the BS first encodes the data symbol of each user according to its pattern vector and then transmits the superimposed encoded symbols $\sum_{j=1}^K \mathbf{v}_j$ through the channel. Therefore, the received signal at the k -th user can be expressed as

$$\mathbf{y}_k = \text{diag}(\mathbf{h}_k) \sum_{j=1}^K \mathbf{g}_j x_j + \mathbf{n}_k = \mathbf{H}_k \mathbf{x} + \mathbf{n}_k, \quad (4)$$

where $\mathbf{H}_k = \text{diag}(\mathbf{h}_k) \mathbf{G}_{M \times K}$ is the PDMA equivalent downlink channel response of the k -th user. Moreover, \mathbf{h}_k and \mathbf{n}_k are, respectively, the downlink channel response and noise vector at the k -th user with length M [16]. We further assume that the channel state information (CSI) is available at the transmitter side such that the channel gains are equalized from the receiver viewpoint. Therefore, the received signal vector for both uplink and downlink transmissions can be expressed as $\mathbf{y} = \mathbf{G} \mathbf{x} + \mathbf{n}$.

As a rather simple multiuser detection algorithm in NOMA systems, SIC properly trades off between the system performance and complexity. However, SIC receivers suffer from error propagation problem meaning that the system performance highly relies on the correctness of early-detected users. To deal with this problem, we can either improve the reliability of the firstly-decoded users or employ more advanced detection algorithms such as ML and MAP. In [16], disparate diversity orders are adopted for different users by assigning patterns with heavier weights to those early-detected users in order to increase

their transmission reliability. Furthermore, employing more advanced detection algorithms severely increases the system complexity especially for larger dimension of the pattern matrix; this may hinder their practical implementation particularly for downlink transmission where the users are supposed to have a lower computational resources than the BS. In the next section, we elaborate how our proposed method can support much larger dimensions of the pattern matrices, even with heavier pattern weights, to boost massive connectivity without a significant increase on the overall system complexity.

III. PROPOSED DESIGN METHOD

A. Fundamentals of the Recursive Pattern Matrix Design

Depending on the number of users and the available REs to them, we propose factorizing the pattern matrix as the Kronecker product of L smaller factor matrices as

$$\mathbf{G}_{M \times K} = \mathbf{G}_{m_1 \times k_1}^{(1)} \otimes \mathbf{G}_{m_2 \times k_2}^{(2)} \otimes \cdots \otimes \mathbf{G}_{m_L \times k_L}^{(L)}, \quad (5)$$

where L is a design parameter, and \otimes denotes the Kronecker product defined as

$$\mathbf{A}_{m \times k} \otimes \mathbf{B}_{m' \times k'} = \begin{bmatrix} a_{11} \mathbf{B} & \cdots & a_{1k} \mathbf{B} \\ \vdots & \ddots & \vdots \\ a_{m1} \mathbf{B} & \cdots & a_{mk} \mathbf{B} \end{bmatrix} \quad (6)$$

for any two matrices \mathbf{A} and \mathbf{B} . Then the dimensions of the resulting pattern matrix in (5) can be obtained based on the dimensions of the factor matrices as $M = \prod_{l=1}^L m_l$, and $K = \prod_{l=1}^L k_l$. In general, if at least one of M or K is not a prime number, we can find some $m_l > 1$ or $k_l > 1$ to factorize the pattern matrix as the Kronecker product of some smaller factor matrices. On the other hand, if both M and K are prime numbers, L is equal to one and the design procedure simplifies to that of conventional pattern matrix design (such as PDMA [16]). Note that both M and K are design parameters and we can always group an appropriate number of users with a specific number of REs to optimize the system in terms of performance and complexity.

The proposed structure not only alleviates the detection complexity at the receiver side (through recursive detection as will be clarified through an illustrating example in the next subsection) but also significantly reduces the search domain at the transmitting party, enabling the usage of large pattern matrices with a reasonable complexity for massive communications. In fact, a regular pattern matrix design requires a comprehensive search over all $\binom{2^M - 1}{K}$ possible $(M \times K)$ -dimensional matrices with distinct nonzero columns (patterns assigned to each user) to find an optimal pattern matrix [16]. On the other hand, it is easy to show that satisfying distinct nonzero columns for the overall pattern matrix $\mathbf{G}_{M \times K}$ of the form (5) requires distinct nonzero columns for all of the factor matrices $\mathbf{G}_{m_l \times k_l}^{(l)}$, $l = 1, 2, \dots, L$. In this case, the total search domain for our proposed design method reduces to $\prod_{l=1}^L \binom{2^{m_l} - 1}{k_l}$ which is much smaller than $\binom{2^M - 1}{K}$. For example, for $M = 6$ and $K = 9$, the regular pattern matrix design requires searching over $\binom{2^6 - 1}{9} = 2.36 \times 10^{10}$ possible matrices, while our design method with factorizing the overall pattern matrix as the Kronecker product of $\mathbf{G}_{2 \times 3}^{(1)} \otimes \mathbf{G}_{3 \times 3}^{(2)}$ only needs to search over $\binom{2^2 - 1}{3} \cdot \binom{2^3 - 1}{3} = 35$ matrices. Note that if, for some l , $2^{m_l} - 1 < k_l$, then we cannot satisfy distinct nonzero columns for the factor matrix $\mathbf{G}_{m_l \times k_l}^{(l)}$. Consequently, as it will be demonstrated in the following example, the overall pattern matrix $\mathbf{G}_{M \times K}$ will have some repeated columns, i.e.,

the same pattern vectors for some of the users. In such circumstances, we can either use more advanced detection methods such as MAP, though over much smaller dimensions, or assign different power coefficients to the users having the same pattern vectors. It is worth mentioning that recursive construction of large matrices based on the Kronecker product of some smaller matrices has been used in different contexts such as polar coding [19] and its later versions like compound polar coding [20].

B. An Illustrating Example

Here we provide an illustrating example to clarify the recursive detection method and facilitate the performance characterization procedures in the next section.

Example 1. Consider a code-domain NOMA network with $K = 18$ users over $M = 9$ REs realized using the Kronecker product of the following three factor matrices

$$\mathbf{G}_{1 \times 2}^{(1)} = \begin{bmatrix} 1 & 1 \end{bmatrix}, \quad \mathbf{G}_{3 \times 3}^{(2)} = \mathbf{G}_{3 \times 3}^{(3)} = \begin{bmatrix} 1 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 1 \end{bmatrix}. \quad (7)$$

Then the resulting overall pattern matrix can be obtained as $\mathbf{G}_{9 \times 18} = \mathbf{G}_{1 \times 2}^{(1)} \otimes \mathbf{G}_{3 \times 3}^{(2)} \otimes \mathbf{G}_{3 \times 3}^{(3)}$. And assuming the availability of CSI, the received signal vector, either in uplink or downlink reception, can be represented as $\mathbf{y}_{9 \times 1} = \mathbf{G}_{9 \times 18} \mathbf{x}_{18 \times 1} + \mathbf{n}_{9 \times 1}$.

It is easy to verify that the k' -th and $(k' + 9)$ -th columns, $k' = 1, 2, \dots, 9$, of the overall pattern matrix $\mathbf{G}_{9 \times 18}$ are the same, meaning that both the k' -th and $(k' + 9)$ -th users have the same pattern vector $\mathbf{g}_{k'}$. This is because both of the elements of $\mathbf{G}_{1 \times 2}^{(1)}$ are the same and this duplicates $\mathbf{G}_{3 \times 3}^{(2)} \otimes \mathbf{G}_{3 \times 3}^{(3)}$ in the construction of the overall pattern matrix. As a result, the data symbols of the k' -th and $(k' + 9)$ -th users, i.e., $x_{k'}$ and $x_{k'+9}$, are always coupled together, leading us to define new symbols $t_{k'} = x_{k'} + x_{k'+9}$. Therefore, we can write the received signals at different REs y_i 's, $i = 1, 2, \dots, 9$, as the linear combination of $t_{k'}$'s, defined based on the rows of the square matrix $\mathbf{G}_{3 \times 3}^{(2)} \otimes \mathbf{G}_{3 \times 3}^{(3)}$, plus the noise components n_i 's (e.g., the received signal at the first RE is given by $y_1 = x_1 + x_2 + x_4 + x_5 + x_{10} + x_{11} + x_{13} + x_{14} + n_1 = t_1 + t_2 + t_4 + t_5 + n_1$). The resulting set of 9 equations for y_i 's can be analyzed using different multiuser detection methods to find $t_{k'}$'s. However, the proposed design method in this paper can recursively find the data symbols with much lower complexity. Here, for the sake of brevity, we focus on the detection of t_1 , t_4 , and t_7 . The rest of the symbols can be detected in a similar recursive procedure. Let us define new variables Z_1 , Z_4 , and Z_7 as the combination of some of the received signals as

$$\begin{aligned} Z_1 &= y_1 + y_2 - y_3 = 2t_1 + 2t_4 + n'_1, \\ Z_4 &= y_4 + y_5 - y_6 = 2t_1 + 2t_7 + n'_4, \\ Z_7 &= y_7 + y_8 - y_9 = 2t_4 + 2t_7 + n'_7, \end{aligned} \quad (8)$$

where $n'_1 = n_1 + n_2 - n_3$, $n'_4 = n_4 + n_5 - n_6$, and $n'_7 = n_7 + n_8 - n_9$. Therefore, the effective signal-to-noise ratio (SNR) is increased by a factor of 4/3 because the noise elements in different REs $n_{k'}$'s are independent, i.e., each of $n_{k'}$'s has the variance of $\sigma'^2 = 3\sigma^2$, where σ^2 is the variance of original noise elements $n_{k'}$'s. Again, we can apply the recursive procedure to get much simpler equations as

$$\begin{aligned} U_1 &= Z_1 + Z_4 - Z_7 = 4t_1 + n''_1, \\ U_4 &= Z_1 + Z_7 - Z_4 = 4t_4 + n''_4, \\ U_7 &= Z_4 + Z_7 - Z_1 = 4t_7 + n''_7, \end{aligned} \quad (9)$$

in which $n''_1 = n'_1 + n'_4 - n'_7$, $n''_4 = n'_1 + n'_7 - n'_4$, and $n''_7 = n'_4 + n'_7 - n'_1$. Therefore, the effective SNR is increased again

by a factor of 4/3 as the resulting noise components $n''_{k'}$'s have the variance of $\sigma''^2 = 3\sigma'^2 = 3^2\sigma^2$. Finally, based on (9), the original data symbols x_k 's can be obtained using optimal MAP detection over simple equations of the form

$$U_{k'} = 4x_{k'} + 4x_{k'+9} + n''_{k'}, \quad k' = 1, 2, \dots, 9 \quad (10)$$

which have much smaller dimensions (i.e., 1×2) compared to the original pattern matrix of dimension 9×18 .

Remark 1. Although the effective reduction on the dimension of equations, using the proposed recursive detection, enables the application of advanced detection methods such as MAP, we can also exploit the power-domain NOMA concept to group the users having the same pattern vector (here the k' -th and $(k' + 9)$ -th users) in order to differentiate between those users and further boost the performance and reduce the complexity of the system; this interesting problem will be explored in our future work.

IV. PERFORMANCE CHARACTERIZATION

The previous example gives a good idea about the potentials of the proposed code-domain NOMA approach and illustrates the process of the low-complexity recursive detection. However, it raises a couple of important questions which should be carefully addressed. In particular, what kind of matrices should be selected and how we can find the optimal factor matrices. How the received signals in different REs can be combined to get smaller dimensions and simpler sets of equations (e.g., converting (8) to (9)). What exactly the gain of such a combining will be in terms of increasing the effective SNR. How many equations and with what dimensions will be left at the end to perform the advanced detection algorithms, such as MAP, and then what would be the overall detection complexity? And, given a set of factor matrices to construct a pattern matrix, what is the overall system performance in terms of the average sum rate? In this section, in order to appropriately answer these issues, we consider a special case where $\mathbf{G}_{m_1 \times k_1}^{(1)} = \mathbf{F}_{m_f \times k_f}$ is a rectangular matrix with $m_f < k_f$ and all of the other $r = L - 1$ factor matrices are the same square matrices as $\mathbf{G}_{m_l \times k_l}^{(l)} = \mathbf{P}_{m_p \times m_p}$, $l = 2, 3, \dots, L$ (hence, the overload factor is $\beta = k_f/m_f > 1$), and leave the general study on the pattern matrix construction to the future research. Therefore, the overall pattern matrix can be represented as $\mathbf{G}_{M \times K} = \mathbf{F}_{m_f \times k_f} \otimes \mathbf{P}_{m_p \times m_p}^{\otimes r}$, where $\mathbf{P}^{\otimes r}$ implies the r -times Kronecker product of \mathbf{P} with itself.

A. Pattern Matrix Design and Recursive Combining

With the considered pattern matrix structure, $\mathbf{G} = \mathbf{F} \otimes \mathbf{P}^{\otimes r}$, both the combining procedure of the received signals in different REs and the resulting gains in the average SNRs, in each recursion, directly relate to the square factor matrix \mathbf{P} . Given the v -th possible factor matrix \mathbf{P}_v , $v = 1, 2, \dots, \binom{2^{m_p}-1}{m_p}$, at each l -th iteration, we have some set of auxiliary equations of the form

$$\begin{bmatrix} y_1^l \\ y_2^l \\ \vdots \\ y_{m_p}^l \end{bmatrix} = \begin{bmatrix} p_{11}^v & p_{12}^v & \cdots & p_{1m_p}^v \\ p_{21}^v & p_{22}^v & \cdots & p_{2m_p}^v \\ \vdots & \vdots & \ddots & \vdots \\ p_{m_p 1}^v & p_{m_p 2}^v & \cdots & p_{m_p m_p}^v \end{bmatrix} \begin{bmatrix} x_1^l \\ x_2^l \\ \vdots \\ x_{m_p}^l \end{bmatrix} + \begin{bmatrix} n_1^l \\ n_2^l \\ \vdots \\ n_{m_p}^l \end{bmatrix}, \quad (11)$$

where p_{ij}^v is the (i, j) -th element of the factor matrix \mathbf{P}_v . And y_j^l , x_j^l , and n_j^l are respectively the j -th auxiliary received signal, data symbol, and noise component at the l -th recursion (see, e.g., the right hand side of (8) and recall that, with some changes of notation, Z_j , $2t_j$, and n'_j are equivalent to y_j^l , x_j^l , and n_j^l in the notation of Eq. (11)).

Algorithm 1 Calculation of the combining coefficients and the corresponding individual SNRs

```

1: Input: dimension of the square matrix  $m_p$ 
2: Output: combining coefficients and SNRs for all  $v$ 's and  $j$ 's

3: for  $v = 1 : \binom{2^{m_p}-1}{m_p}$  do
4:    $\mathbf{P}_v = [p_{ij}^v]_{m_p \times m_p}$ 
5:   for  $j = 1 : m_p$  do
6:     Find all sets of coefficients  $\{\alpha_{i=1:m_p}^{(v,j)}\}$  Such That:
7:       C1:  $\alpha_i^{(v,j)} \in \{-1, 0, 1\}$ 
8:       C2:  $\sum_{i=1}^{m_p} \alpha_i^{(v,j)} p_{ij}^v = w'_{v,j} \neq 0$ 
9:       C3:  $\sum_{i=1}^{m_p} \alpha_i^{(v,j)} p_{ij'}^v = 0, j' \neq j = 1, 2, \dots, m_p$ 
10:    Calculate  $\gamma'_{v,j} = [w'_{v,j}]^2 / \sum_{i=1}^{m_p} (\alpha_i^{(v,j)})^2$ 
11:    Output  $\{\alpha_{i=1:m_p}^{(v,j)}\} = \text{argmax}_{\{\alpha_{i=1:m_p}^{(v,j)}\}} \gamma'_{v,j}$  over  $\{\alpha_{i=1:m_p}^{(v,j)}\}$ 
12:    Output  $\gamma_{v,j} = [\sum_{i=1}^{m_p} \alpha_i^{(v,j)} p_{ij}^v]^2 / \sum_{i=1}^{m_p} (\alpha_i^{(v,j)})^2$ 
13:   end for
14: end for

```

In order to obtain the effective combining coefficients (such as combining Z_j 's in (9)) that facilitate the recursive detection with a low complexity and in the meantime provide the maximum increase of the SNRs, we find the so-called systematic representation of \mathbf{P}_v , i.e., for each j -th data symbol we find all the linear combinations of the rows of the factor matrix that result in the form $x_j^l \sum_{i=1}^{m_p} \alpha_i^{(v,j)} p_{ij}^v$ which only contains x_j^l . In this representation, $\alpha_i^{(v,j)} \in \{-1, 0, 1\}$, $i = 1, 2, \dots, m_p$, are the combining coefficients. Such a combination, due to the independence of the noise components n_j^l 's, multiplies the SNR of the previous iteration by $\gamma_{v,j} = w_{v,j}^2 / \sum_{i=1}^{m_p} (\alpha_i^{(v,j)})^2$, where $w_{v,j} = \sum_{i=1}^{m_p} \alpha_i^{(v,j)} p_{ij}^v$ is the weight of the desired data symbol x_j^l after the combination. Then the selection procedure for the combining coefficients that yield the maximum SNR for each auxiliary data symbol can be summarized as Algorithm 1.

Remark 2. The optimization procedure in Algorithm 1 is performed over all $\binom{2^{m_p}-1}{m_p}$ possible factor matrices with distinct nonzero columns, and then the optimal square factor matrix, the gains on the SNRs, and the efficient combining coefficients are obtained by choosing the best answer set satisfying further constraints, e.g., maximizing the average sum rate. We should emphasize based on our hierarchical construction that even for the pattern matrices with large overall dimensions the search process in Algorithm 1 is performed over square factor matrices \mathbf{P} with much smaller dimensions m_p ; this further paves the way toward massive code-domain NOMA.

Remark 3. It is easy to verify that for the factor matrices with linearly-independent rows there is only one set of coefficients satisfying the constraints in Algorithm 1 for each data symbol.

B. Detection Procedure and its Complexity

In this subsection, we first elaborate the proposed recursive detection method and then characterize its complexity.

1) *Detection Procedure:* Assume that the communication protocol is established between the transmitters and receivers, i.e., the design parameters such as the optimal square factor matrix $\mathbf{P}_{m_p \times m_p}$ and combining coefficients from Algorithm 1 (also recall Remark 2), number of recursions $r = L - 1$, and the rectangular matrix $\mathbf{F}_{m_f \times k_f}$ are known for the receiver (either a typical user during the downlink phase or the BS during

the uplink transmission); therefore, the pattern matrix structure $\mathbf{G}_{M \times K} = \mathbf{F}_{m_f \times k_f} \otimes \mathbf{P}_{m_p \times m_p}^{\otimes r}$ is known. Given the combining coefficients, we can further define an $m_p \times m_p$ matrix α comprising the combining vectors $\{\alpha_i^{(j)}\}_{i=1}^{m_p}$ as its j -th row such that the matrix product of $\alpha \mathbf{P}$ gives an $m_p \times m_p$ diagonal matrix with the j -th diagonal element $w_j = \sum_{i=1}^{m_p} \alpha_i^{(j)} p_{ij}$ being the weight of the j -th data symbol after the combination (recall the constraints for the combining coefficients from Algorithm 1).

First Recursion: In order to perform the detection, the receiver in the first recursion takes the vector of received signals over $M_1 = M = m_f m_p^r$ REs, i.e., y_i 's, $i = 1, 2, \dots, M$, and divides them into $M_1/m_p = m_f m_p^{r-1}$ groups of m_p equations; therefore, the i_1 -th group includes $\{y_{(i_1-1)m_p+1}, y_{(i_1-1)m_p+2}, \dots, y_{i_1 m_p}\}$. Each of these y_i 's contains at most K different transmitted symbols x_k 's, $k = 1, 2, \dots, K$, since each one is constructed as a linear combination of x_k 's, defined based on the i -th row of the overall pattern matrix $\mathbf{G}_{M \times K}$, and the noise component n_i over the i -th RE. The receiver then combines the elements of each group using the combining matrix α to form new m_p symbols at each group; the first new symbol is constructed by combining the previous symbols using the first row of α and so on. Note that such a combination can be expressed in a matrix form through multiplying α by a new matrix comprising the m_p rows of \mathbf{G} corresponding to the symbols of each group. Then it is easy to verify that the m' -th new equation of each group (which is constructed through the m' -th row of α), $m' = 1, 2, \dots, m_p$, contains at most $K_1 = K/m_p = k_f m_p^{r-1}$ different symbols from the set $\{x_{m'}, x_{m'+m_p}, \dots, x_{m'+k_f m_p^{r-1}-m_p}\}$ (recall that the product of $\alpha \mathbf{P}$ is a diagonal matrix), i.e., the number of unknown variables is reduced by a factor of m_p . Also the effective SNR of each symbol in the m' -th new equation is increased by a factor of $\gamma_{m'}$.

Second Recursion: Note based on the above explanation that after the first recursion the m' -th new equation of each group can only contain symbols from the set $\{x_{m'}, x_{m'+m_p}, \dots, x_{m'+k_f m_p^{r-1}-m_p}\}$. This means different equations of a given group contain disjoint sets of symbols while equations with the same index of different groups (e.g., the m' -th equation of all groups) contain symbols from the same set. Therefore, the receiver in the second recursion forms m_p super-groups of $M_2 = M_1/m_p = m_f m_p^{r-1}$ equations (note that in the first recursion we had $m_p^0 = 1$ super-group of size $M_1 = M$) by placing the m' -th new equation of each of those $m_f m_p^{r-1}$ groups in the first recursion into the m' -th super-group. Now, the receiver follows exactly the same procedure as the first recursion over each of these disjoint m_p super-groups with the m_p times smaller size of M_2 , i.e., divides the equations in each of super-groups into $M_2/m_p = m_f m_p^{r-2}$ groups of m_p equations and combines the signals within each group using the combining matrix α . Following the same logics we can argue that the maximum number of unknown variables at each of the new equations is reduced by a factor of m_p from K_1 to $K_2 = K_1/m_p = k_f m_p^{r-2}$, and the SNR of the symbols in the m'' -th new equation, $m'' = 1, 2, \dots, m_p$, of each of the groups in the m' -th super-group is increased by a factor of $\gamma_{m''}$ from $\gamma_{m'}$ in the first recursion to $\gamma_{m'} \gamma_{m''}$.

Final Recursion: By induction, it is easy to verify that in the r -th recursion we will have m_p^{r-1} super-groups of size $M_r = M/m_p^{r-1} = m_f m_p$. Therefore, the receiver in the r -th iteration divides the equations within each of these M_r super-groups into $M_r/m_p = m_f$ groups of size m_p and combines the symbols of each group using α to get equations containing at most

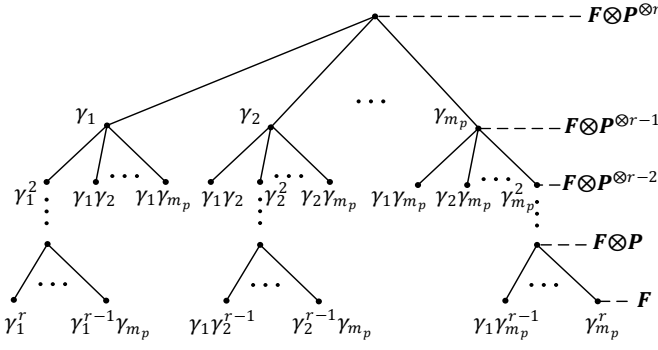


Figure 1. Tree diagram representation of the recursive detection and the resulting individual SNR gains.

$K_r = K/m_p^r = k_f$ unknown variables. Based on these arguments and following the same logics we can conclude that after the r -th iteration we end up with m_p^r sets of m_f equations each containing at most $K_r = k_f$ unknown variables which are defined based on the form of the rectangular matrix \mathbf{F} and can be processed using advanced multiuser detection algorithms such as MAP though over the much smaller dimension of $m_f \times k_f$ instead of $M \times K$. The recursive detection procedure is schematically shown through a tree diagram in Fig. 1 which further helps to understand the progressive development of the SNRs after each recursion.

2) *Detection Complexity*: Based on the above comprehensive arguments, we can now characterize the detection complexity of the proposed recursive algorithm to make sure its applicability. Note that our recursive procedure results in a relatively small size for each factor matrix, even for a large M and K ; hence, it is more informative to obtain the bounds on the number of required operations instead of their orders which assumes the dimension of matrices grow large.

Proposition 1. For the proposed pattern matrix structure $\mathbf{G} = \mathbf{F} \otimes \mathbf{P}^{\otimes r}$, the total number of additions/subtractions and multiplications required for the recursive detection are at most

$$\begin{aligned} \mathcal{N}_{\text{add}}^{\text{rcv}}(m_f, m_p, r) &= r m_f m_p^r (m_p - 1) + m_p^r \mathcal{N}_{\text{add}}^{\text{rgl}}(\mathbf{F}_{m_f \times k_f}), \\ \mathcal{N}_{\text{mul}}^{\text{rcv}}(m_f, m_p, r) &= m_p^r \mathcal{N}_{\text{mul}}^{\text{rgl}}(\mathbf{F}_{m_f \times k_f}), \end{aligned} \quad (12)$$

respectively, where $\mathcal{N}_{\text{add}}^{\text{rgl}}(\mathbf{F}_{m_f \times k_f})$ and $\mathcal{N}_{\text{mul}}^{\text{rgl}}(\mathbf{F}_{m_f \times k_f})$ are the number of additions/subtractions and multiplications in a regular PDMA system [16] that can be obtained given the constellation size, characteristics of the pattern matrix $\mathbf{F}_{m_f \times k_f}$, and the detection method applied to \mathbf{F} such as MAP, ML, and MPA.

Proof: Based on the detection procedure in Section IV-B1, for the pattern matrix structure $\mathbf{G}_{M \times K} = \mathbf{F}_{m_f \times k_f} \otimes \mathbf{P}_{m_p \times m_p}^{\otimes r}$, we will end up with $M = m_f m_p^r$ equations representing the received signals over each of M REs. By applying our recursive combining method to these M equations we can convert them to m_p^r set of m_f equations, each containing at most k_f auxiliary data symbols, defined based on the rectangular factor matrix \mathbf{F} . In order to obtain each of these M new equations, we need to combine at most m_p equations, using $m_p - 1$ additions/subtractions, at each of r recursions. Therefore, these new m_p^r sets of m_f equations can be obtained using only at most $r m_f m_p^r (m_p - 1)$ additions/subtractions. Then we can apply any advanced detection method to the resulting m_p^r sets of equations which requires an additional m_p^r times the number of operations over a much smaller factor matrix \mathbf{F} . ■

Example 2. A PDMA system with the considered parameters in Example 1 needs only 9 times the required operations for applying any advanced detection algorithm to a regular PDMA

system with a much smaller pattern matrix $\mathbf{F} = [1 \ 1]$, and an extra 36 additions/subtractions.

Remark 4. The proposed method can also significantly reduce the latency because it enables parallel processing for the multiuser detection at the receiver side. In fact, as shown in Example 1 and Fig. 1, we can apply the recursive detection to simultaneously process the data of different users at each recursion, i.e., concurrently process m_p^{l-1} sets of equations (super-groups) at each l -th recursion, $l = 1, 2, \dots, r$. It is worth noting that for the regular PDMA design, the rather simple SIC detection needs to process the data of each user one-by-one in a serial fashion.

C. Average Sum Rate

Based on the discussions in the previous parts and the system model in this paper, we can obtain the average sum rate of the system using the following theorem.

Theorem 2. The per-RE average sum rate of code-domain NOMA with the pattern matrix $\mathbf{G}_{M \times K} = \mathbf{F}_{m_f \times k_f} \otimes \mathbf{P}_{m_p \times m_p}^{\otimes r}$ and the proposed recursive detection can be expressed as

$$\begin{aligned} C_M &= \frac{1}{2M} \sum_{r_1 + r_2 + \dots + r_{m_p} = r} \frac{r!}{r_1! r_2! \dots r_{m_p}!} \\ &\quad \times \log_2 \det \left(\mathbf{I}_{m_f} + \rho \gamma_1^{r_1} \gamma_2^{r_2} \dots \gamma_{m_p}^{r_{m_p}} \mathbf{F} \mathbf{F}^t \right), \end{aligned} \quad (13)$$

where \mathbf{I}_{m_f} is an $m_f \times m_f$ identity matrix, \mathbf{F}^t is the transpose of \mathbf{F} , and $\rho = P_x / \sigma^2$ in which $P_x = \mathbb{E}[x_j^2]$, $j = 1, 2, \dots, K$, and $\sigma^2 = \mathbb{E}[n_i^2]$, $i = 1, 2, \dots, M$.

Proof Sketch: It is well known that for a regular PDMA with the pattern matrix $\mathbf{A}_{M \times K}$ and the average SNR of ρ for all of the received original data symbols, the per-RE average sum rate is given by $C_M^{\text{PDMA}} = \frac{1}{2M} \log_2 \det(\mathbf{I}_M + \rho \mathbf{A} \mathbf{A}^t)$ for the optimal MAP detection [18]. Using the recursive detection, as elaborated in Section IV-B1 and schematically shown using a tree diagram in Fig. 1, each of r square matrices manifest their effect as an increase on the SNR and finally we end up with a rectangular factor matrix \mathbf{F} . Then it is easy to verify using the tree diagram that square matrices increase the average SNR as a factor of $\gamma_1^{r_1} \gamma_2^{r_2} \dots \gamma_{m_p}^{r_{m_p}}$ such that $r_1 + r_2 + \dots + r_{m_p} = r$. ■

Remark 5. Note that $\sum_{\Lambda_r} \frac{r!}{r_1! r_2! \dots r_{m_p}!} = m_p^r$, where $\Lambda_r = \{r_1, r_2, \dots, r_{m_p} : \sum_{i=1}^{m_p} r_i = r\}$. Therefore, (13) contains the summation of m_p^r REs embedded in the Kronecker product of $\mathbf{P}^{\otimes r}$, and the remaining portion of $m_f = M/m_p^r$ will be spanned by the rectangular matrix \mathbf{F} in (13).

Example 3. For the system parameters in Example 1 the average sum rate of all users per RE in (13) simplifies to $C_9 = 0.5 \log_2(1 + 2(4/3)^2 \rho)$, which can also be inferred from (10).

Remark 6; Optimal Factor Matrices. When the design target is to maximize the average sum rate, the *optimal pattern matrix* design can be formulated as the selection of an appropriate rectangular matrix \mathbf{F}^* together with the v^* -th square matrix in Algorithm 1 with the corresponding set of SNRs $\{\gamma_{v^*,1}, \gamma_{v^*,2}, \dots, \gamma_{v^*,m_p}\}$, such that they result in the maximum sum rate in (13). Satisfying further design targets such as the users' individual rate [21] deserves future studies.

Remark 7. Given the v -th square factor matrix \mathbf{P}_v , we can apply SIC detection to some of the m_p equations in (11) to detect the j -th auxiliary data symbol x_j^l with a higher SNR $\gamma_{v,j}$, for some $j \in \{1, 2, \dots, m_p\}$. This evidently increases the average sum rate in (13); however, it may increase the error probability due to the error propagation problem in SIC detection, resulting in an interesting trade-off between the rate and reliability. This mechanism also increases the latency because we should wait for the detection of some symbols before starting the detection of the others.

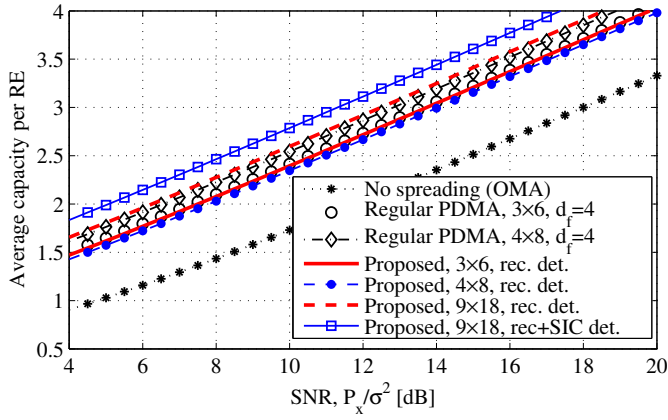


Figure 2. Average capacity of various multiple access mechanisms including the traditional OMA, regular PDMA, and our proposed recursive method.

Example 4. After detecting the auxiliary data symbols t_1 and t_4 in (8) using the first and second equations in (9), we can form $Z_4 + Z_7 - t_1 - t_4 = 4t_7 + n'_4 + n'_7$, instead of the third equation of (9), to detect t_7 with a twice larger SNR rather than the factor of $4/3$. In this case, using (13), the average sum rate will increase to $C_9 = (4/18) \log_2(1 + 2(4/3)^2 \rho) + (4/18) \log_2(1 + 2(8/3)\rho) + (1/18) \log_2(1 + 2(2)^2 \rho)$.

V. NUMERICAL RESULTS

In this section, we provide the numerical results for the average sum rate of various configurations, including the traditional OMA, regular PDMA with the reported optimal pattern matrices in [16] such as 3×6 and 4×8 matrices with the row weight $d_f = 4$, and our proposed recursive code-domain NOMA with the square factor matrices obtained from Algorithm 1. By running Algorithm 1 for $m_p = 3$, we get the optimal square matrix $\mathbf{G}_{3 \times 3}$ in Eq. (7) and the combining coefficients and individual SNRs reported in Example 1. Moreover, for $m_p = 4$ Algorithm 1 results in four optimal square matrices of the same form and the same set of SNRs. A sample of such an optimal square matrix $\mathbf{P}_{4 \times 4}$ along with the corresponding combining matrix $\alpha_{4 \times 4}$ is as follows

$$\mathbf{P}_{4 \times 4} = \begin{bmatrix} 0 & 0 & 0 & 1 \\ 0 & 1 & 1 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 1 & 0 & 0 \end{bmatrix}, \quad \alpha_{4 \times 4} = \begin{bmatrix} 0 & -1 & 1 & 1 \\ 0 & 1 & -1 & 1 \\ 0 & 1 & 1 & -1 \\ 1 & 0 & 0 & 0 \end{bmatrix}. \quad (14)$$

Therefore, based on the combinations defined by the matrix product of $\alpha \mathbf{P}$, we get the SNR gains of $\gamma_1 = \gamma_2 = \gamma_3 = 4/3$, and $\gamma_4 = 1$.

Fig. 2 shows the per-RE average sum rate of various multiple access techniques. It is observed that both regular PDMA and our proposed code-domain NOMA mechanisms significantly outperform the traditional OMA method. For the simulation of our proposed method we obtained the square factor matrices from Algorithm 1 and assumed the rectangular factor matrix $\mathbf{F} = [1 \ 1]$. Fig. 2 demonstrates that, for relatively similar system parameters, our proposed method with the simple recursive detection approaches the capacity of regular PDMA with the optimal pattern matrix $\mathbf{A}_{M \times K}$ and optimal MAP detection, i.e., $C_M^{\text{PDMA}} = \frac{1}{2M} \log_2 \det(\mathbf{I}_M + \rho \mathbf{A} \mathbf{A}^t)$. However, due to the remarkably lower complexity of our method, we can exploit higher pattern matrix dimensions (by increasing r) and also apply SIC detection at some recursions similar to Example 3 to get larger sum rates while achieving such gains through larger pattern matrix dimensions are questionable for the regular PDMA.

VI. CONCLUSION

In this paper, we proposed a low-complexity recursive approach toward code-domain NOMA based on factorizing the spreading pattern matrix as the Kronecker product of some smaller factor matrices. We observed that the proposed method, with the simple recursive detection, approaches the average sum rate of regular PDMA technique with optimal detection. The proposed method benefits from a significantly lower complexity and latency and, through realizing large-dimension pattern matrices, paves the way toward massive nonorthogonal multiplexing for the broad range of 5G applications.

REFERENCES

- [1] L. Dai, B. Wang, Y. Yuan, S. Han, I. Chih-Lin, and Z. Wang, "Non-orthogonal multiple access for 5G: solutions, challenges, opportunities, and future research trends," *IEEE Commun. Mag.*, vol. 53, no. 9, pp. 74–81, 2015.
- [2] Y. Chen, A. Bayesteh, Y. Wu, B. Ren, S. Kang, S. Sun, Q. Xiong, C. Qian, B. Yu, Z. Ding *et al.*, "Toward the standardization of non-orthogonal multiple access for next generation wireless networks," *IEEE Commun. Mag.*, vol. 56, no. 3, pp. 19–27, 2018.
- [3] Z. Ding, X. Lei, G. K. Karagiannis, R. Schober, J. Yuan, and V. K. Bhargava, "A survey on non-orthogonal multiple access for 5G networks: Research challenges and future trends," *IEEE J. Sel. Areas Commun.*, vol. 35, no. 10, pp. 2181–2195, 2017.
- [4] Z. Ding, Z. Yang, P. Fan, and H. V. Poor, "On the performance of non-orthogonal multiple access in 5G systems with randomly deployed users," *IEEE Signal Process. Lett.*, vol. 21, no. 12, pp. 1501–1505, 2014.
- [5] Z. Ding, M. Peng, and H. V. Poor, "Cooperative non-orthogonal multiple access in 5G systems," *IEEE Commun. Lett.*, vol. 19, no. 8, pp. 1462–1465, 2015.
- [6] Y. Liu, Z. Ding, M. Elkashlan, and H. V. Poor, "Cooperative non-orthogonal multiple access with simultaneous wireless information and power transfer," *IEEE J. Sel. Areas Commun.*, vol. 34, no. 4, pp. 938–953, 2016.
- [7] Q. Sun, S. Han, I. Chin-Lin, and Z. Pan, "On the ergodic capacity of MIMO NOMA systems," *IEEE Wireless Commun. Lett.*, vol. 4, no. 4, pp. 405–408, 2015.
- [8] Z. Ding, P. Fan, and H. V. Poor, "Random beamforming in millimeter-wave NOMA networks," *IEEE Access*, vol. 5, pp. 7667–7681, 2017.
- [9] H. Pan, L. Lu, and S. C. Liew, "Practical power-balanced non-orthogonal multiple access," *IEEE J. Sel. Areas Commun.*, vol. 35, no. 10, pp. 2312–2327, 2017.
- [10] H. Nikopour and H. Baligh, "Sparse code multiple access," in *Proc. IEEE Int. Symp. Pers. Indoor Mobile Radio Commun., London, U.K., Sep. 2013*, pp. 332–336.
- [11] F. Wei and W. Chen, "A low complexity SCMA decoder based on list sphere decoding," in *Global Communications Conference (GLOBECOM), 2016 IEEE*. IEEE, 2016, pp. 1–6.
- [12] D. Fang, Y.-C. Huang, Z. Ding, G. Geraci, S.-L. Shieh, and H. Claussen, "Lattice partition multiple access: A new method of downlink non-orthogonal multiuser transmissions," in *Global Communications Conference (GLOBECOM), 2016 IEEE*. IEEE, 2016, pp. 1–6.
- [13] M. Qiu, Y.-C. Huang, S.-L. Shieh, and J. Yuan, "A lattice-partition framework of downlink non-orthogonal multiple access without SIC," *Accepted for Publication in IEEE Trans. Commun.*, 2018.
- [14] Q. Xiong, C. Qian, B. Yu, and C. Sun, "Advanced NoMA scheme for 5G cellular network: Interleave-grid multiple access," in *Globecom Workshops (GC Wkshps), 2017 IEEE*. IEEE, 2017, pp. 1–5.
- [15] X. Dai, S. Chen, S. Sun, S. Kang, Y. Wang, Z. Shen, and J. Xu, "Successive interference cancellation amenable multiple access (SAMA) for future wireless communications," in *Communication Systems (ICCS), 2014 IEEE International Conference on*. IEEE, 2014, pp. 222–226.
- [16] S. Chen, B. Ren, Q. Gao, S. Kang, S. Sun, and K. Niu, "Pattern division multiple access—a novel nonorthogonal multiple access for fifth-generation radio networks," *IEEE Trans. Veh. Technol.*, vol. 66, no. 4, pp. 3185–3196, 2017.
- [17] Z. Yang, J. Cui, X. Lei, Z. Ding, P. Fan, and D. Chen, "Impact of factor graph on average sum rate for uplink sparse code multiple access systems," *IEEE Access*, vol. 4, pp. 6585–6590, 2016.
- [18] O. Shental, B. M. Zaidel, and S. Shamai, "Low-density code-domain NOMA: Better be regular," in *Proc. IEEE Int. Symp. Inf. Theory, Aachen, Germany, Jul. 2017*, pp. 2628–2632.
- [19] E. Arikan, "Channel polarization: A method for constructing capacity-achieving codes for symmetric binary-input memoryless channels," *IEEE Trans. Inf. Theory*, vol. 55, no. 7, pp. 3051–3073, 2009.
- [20] H. Mahdavi, M. El-Khamy, J. Lee, and I. Kang, "Compound polar codes," in *Information Theory and Applications Workshop (ITA)*. IEEE, 2013, pp. 1–6.
- [21] P. Xu, Z. Ding, X. Dai, and H. V. Poor, "A new evaluation criterion for non-orthogonal multiple access in 5G software defined networks," *IEEE Access*, vol. 3, pp. 1633–1639, 2015.